4.1 FUNCTIONAL MODEL

The OAIS of Figure 2-1 is broken into six functional entities and related interfaces as shown in **Figure 4-1**. The lines connecting entities identify communication paths over which information flows in both directions. The lines shown to Administration are dashed only to reduce diagram clutter.

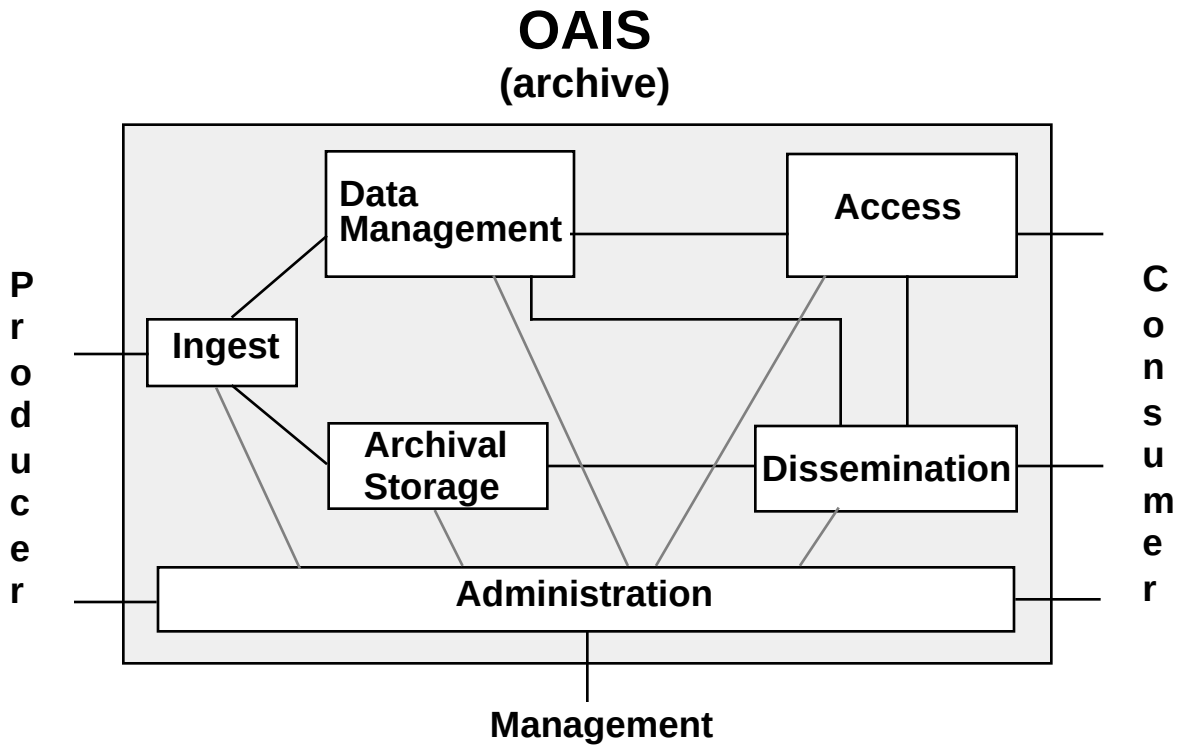

Figure 4-1. OAIS Functional Entities

The role provided by each of the entities in Figure 4-1 is briefly described as follows:

- **Ingest:** This entity provides the services and functions to accept and validate a Submission Information Package (SIP) from Producers and prepare the contents for storage and management within the archive. In summary there is a scheduling function to negotiate a submission agreement and the delivery of one or more SIPs; each SIP is physically received by the staging function; the Content Information (CI) may go through a conversion process to comply with the internal archive data model and descriptive information may be added to enhance the utility of the package and result in an Archival Information Package (AIP); the AIP is reviewed by the archive staff and others; **Descriptive Information** is derived from the AIP for cataloging in the data management system; and the AIP is made available for transfer to Archival Storage.

- **Archival Storage:** This entity provides the services and functions for the storage and retrieval of AIPs. Archival Storage functions include receiving AIPs from staging storage to permanent Archival Storage; managing the Archival Storage hierarchy; physically migrating data to new media over time; performing routine and special error checking; providing backup procedures; and providing access to AIPs for dissemination.

- **Data Management:** This entity provides the services and functions for populating, maintaining, and querying a wide variety of descriptive information such as catalogs, processing histories and processing algorithms. It also provides Ingest, Access, Dissemination and Administration with system information (e.g. consumer access information, security information, and operational

schedules). Data Management functions include providing services for requesting and generating reports; providing the capability for both transactional updates (loading new descriptive information or archive operational statistics) to the data base and periodic review updates; and general data base administration functions (maintaining schema and view definitions and referential integrity). This entity, together with Archival Storage, conceptually contains all the persistent information needed for OAIS operations.

- **Administration:** This entity manages all of the system activities. The Administration functions include planning and scheduling archive facility resources, maintaining configuration management of system hardware and software, performing accounting functions to bill consumers for services; providing service functions to consumers; performing data engineering work to develop and maintain archive standards and policies; and providing a point of interaction with OAIS Management.

- **Access:** This entity supports consumers in determining the existence, description, location and availability of information stored in the OAIS. The Access entity includes functions for providing a mechanism (access session) for the Consumer to communicate with the OAIS; applying access controls to this communication interface; allowing the Consumer to peruse finding aids; and for selecting data objects for on-line or off-line dissemination.

- **Dissemination:** This entity includes the functions which provide archive products to Consumers. The Dissemination entity includes functions for receiving dissemination requests from the Access entity; interacting with Archival Storage to retrieve requested AIPs; generating any necessary descriptive information required to accompany the Dissemination Information Packages (DIP); performing processing that is required to convert the AIP into a DIP; making the resulting DIPs available on-line or off-line; and monitoring the status of the dissemination request until successful completion.

In addition to the entities described above, there are various common services assumed to be available, and all of these are discussed in more detail in Sections 4.1.1 through 4.1.7. Specific flows of information among the entities are shown here in bold type, followed by an identifier in braces. For example, Section 4.1.3 identifies the flow "**storage confirmation {4.1.3g}.**" This and all other flows are then illustrated in the set of data flow diagrams collected in Section 4.1.8.

4.1.1 COMMON SERVICES

Modern, distributed computing applications assume a number of supporting services such as inter-process communication, name services, temporary storage allocation, exception handling, security, and directory services. Common Services provides a single conceptual source for these services.

4.1.2 INGEST

The functions of the Ingest entity are detailed below:

* The **Scheduling** function ~~develops a **submission agreement {4.1.2c}** and~~ negotiates a **data submission schedule {4.1.2k}** with the producer. It also maintains a calendar of expected Data Delivery Sessions that will be needed to transfer one or more complete Archive Information Packages to the OAIS and the resource requirements to support their ingestion.

* The **Staging** function provides the appropriate storage capacity or devices to receive a **SIP {4.1.2a}** from the producer. The SIPs may be delivered via electronic transfer (e.g. ftp); loaded from media submitted to the archive; or simply mounted (e.g. CD-ROM) on the archive file system for access. The Staging function may represent a legal transfer of custody for the CI in the SIP, and may require that special access controls be placed on the contents. The Staging function provides a **confirmation of receipt {4.1.2f}** of a SIP to the Producer.

* The **Conversion** function transforms one or more SIPs into one or more AIPs that conforms to the internal data model of the archive. This may involve file format conversions, data representation conversions or reorganization of the content information in the SIPs.

* The **Review** function provides a validation of the SIP (for a partial ingestion) or the AIP after it has been converted to conform to the internal data model of the archive. The Review function may be carried out on a single SIP or may be withheld until a number of Data Delivery Sessions have been concluded and all the components of the AIP have been accumulated. The review is carried out by the archive data engineers and may also involve an outside committee (e.g., peer review). The review process must verify that the SIP has been physically transported correctly to the archive staging area and successfully converted to internal archive format; that the quality of the data meets the requirements of the archive and the review committee; that there is adequate Representation Information and Preservation Description Information to ensure the Content Information is understandable and independently usable to the Designated Community; and that the Descriptive Information is sufficient to make the preserved information adequately findable by Consumers from the Designated Community. The formality of the review will vary depending on internal archive policies. The review process may determine that some portions of the SIP are not appropriate for inclusion in the archive and must be resubmitted or excluded. After the review process is completed any **liens {4.1.2b}** are reported to the producer who will then **resubmit {4.1.2i}** or **appeal {4.1.2j}** the decision. After the review is completed a **final ingest report {4.1.2g}** on the Data Delivery Session is prepared for distribution to Administration and to the Producer.

* The **Transfer Initiation** function starts the process that moves the AIP from the staging area to the storage area. This may be either an electronic, physical, or a virtual (i.e. data stays in place) transfer. Transfer of the **AIP {4.1.2p}** includes a **storage request {4.1.2h}**. After completing the transfer, Archival Storage returns a **storage confirmation {4.1.2m}** indicating the storage location of the AIP. This confirmation is included in the database update prepared by the Cataloging function.

* The **Cataloging** function extracts descriptive components of the SIPs or completed AIPs to populate the data management system. This **descriptive information {4.1.2n}** consists of selected parameters that are needed to support the functions of the Access and Dissemination entities. The transfer of descriptive information to the Data Management entity includes a **database update request {4.1.2d}.** In return, Data Management provides a **database update response {4.1.2e}** indicating the status of the update.

4.1.3 ARCHIVAL STORAGE

The functions of the Archival Storage entity are detailed below:

* The **Transfer Receiving** function receives a **transfer request {4.1.3d}** of an archival information package from staging Archival Storage and moves the data to permanent Archival Storage within the archive. The transfer request may need to indicate the anticipated frequency of utilization of the data objects comprising the Archival Information Package (AIP) to allow the appropriate storage devices or media to be selected for storing the AIP. The Transfer Receiving function will select the media type, prepare the devices or volumes and perform the physical transfer to the Archival Storage volumes. On completion of the transfer, this function sends an **Archival Storage confirmation {4.1.3g}** message to Ingest.

* The **Hierarchy Management** function positions the AIPs (AIUs and AICs) on the appropriate media based on directions from ingest (transfer request), administrative **policies {4.1.3h}** or usage statistics. This function provides regular reports to Administration summarizing the **inventory of media on-hand {4.1.3a}**, **available storage capacity {4.1.3b}** in the various tiers of the storage hierarchy, and other **operational statistics {4.1.3c}**.

* The **Physical Migration** function provides the capability to reproduce the AIPs over time. Refer to Section 5 for a detailed description of migration issues. The fundamental rule of data migration under the Physical Migration function is that the Content Information and Preservation Description Information must not be altered. However the data constituiting the Packaging Information may be changed as long as it continues to perform the same function. The migration strategy must take into consideration the expected and actual rates of errors encountered in various media types, their performance, and the costs of ownership when deciding what media to migrate to. If media dependent attributes (e.g. tape block sizes, CD-ROM volume information) have been included as part of the Content Information, a way must be found to preserve this information when migrating to higher capacity media with different storage architectures.

* The **Error Checking** function provides statistically acceptable assurance that no components of the archive information package are corrupted during transfer receiving, migration, backup or duplication procedures. This function requires that all hardware and software within the archive provide notification of potential errors and that these errors are routed to standard logs that are checked by the Archival Storage staff. The PDI Fixity Information provides some assurance that the Content Information has not been altered as the AIP is moved and acessed. Similar information is needed to protect the PDI itself. A standard mechanism for tracking and verifying the validity of all data objects within the archive could also be used. For example, cyclical redundancy checks (CRCs) could be maintained for every individual data file. The storage facility procedures should provide for random verification of the integrity of data objects using CRCs or some other error checking mechanism.

* The **Disaster Recovery** function provides a mechanism for producing duplicate copies of AIPs (AIUs and AICs) in the archive collection. The backup media should be capable of being removed from the archive for storage at a separate facility. **Disaster Recovery policies {4.1.3i}** are specified by Administration.

* The **Provide Data** function provides copies of stored AIPs to dissemination. The function receives a **data request {4.1.3j}** from dissemination which identifies the requested **AIP(s) {4.1.3m}** and either the output media type or a staging area for electronic transfers. The Provide Data function sends a **notice of data transfer {4.1.3k}** to Dissemination.

4.1.4 DATA MANAGEMENT

The functions of the Data Management entity are detailed below:

* The **Data Base Administration** function is responsible for maintaining the integrity of the Data Management persistent storage; for creating any schema or table definitions required to support data management functions; and for providing the capability to create, maintain and access customized user views of the contents of this storage.

* The **Report Generation** function receives a **report request {4.1.4b}** from Ingest, Access, Administration or Dissemination and prepares the necessary queries to generate the report.

* The **Report Production** function sends the **report {4.1.4a}** to the requester. It also provides the capability to store report requests and to generate periodic reports or reports triggered by logical criteria on a periodic basis.

* The **Update** function adds, modifies or deletes information in the Data Management persistent storage. There are three major sources of updates; ingest transactions, system updates, and review updates. Ingest transactions identify new AIPs stored in the archive {flow from Ingest}. System updates include all system related information (Consumer information, request tracking). Review updates are generated by periodic reviewing and updating of information values (e.g. contact names, and addresses). The update function provides regular reports to administration summarizing the **status of updates to the data base {4.1.4e}**.

* The **Provide Subscription Requests** function maintains a record of subscription requests and periodically compares it to the contents of the archive to determine if new data is available. If new data is available this function generates a **dissemination request** which is sent to the dissemination function.
~~* The **Data Base Administration** function is responsible for maintaining the integrity of the Data Management persistent storage; for creating any schema or table definitions required to support data management functions; and for providing the capability to create, maintain and access customized user views of the contents of this storage.~~

4.1.5 ADMINISTRATION

The functions of the Administration entity are detailed below:

* The **Planning and Scheduling** function schedules system usage. It keeps records on times of heavy resource utilization, system down times for maintenance, and system upgrades. The planning and scheduling function also **solicits desirable archivable information {4.1.5b}** for inclusion into the OAIS and handles administrative aspects of acquiring new SIPs.

* The **Configuration Management** function maintains configuration control over the archive system, systematically controlling changes to the configuration. This function maintains integrity and tractability of the configuration during all phases of the system life cycle. It also audits system operations, system performance, and system usage.

* The **Physical Access Control** function provides mechanisms to restrict or allow physical access (doors, locks, guards) to elements of the archive as determined by archive policies.

* The **Customer Service** function provides any necessary assistance to archive system consumers. This service will include **answering questions, resolving consumer problems {4.1.5d}**, providing information about and **documentation {4.1.5e}** on the system, providing status of orders, and providing information about the **status of data ingest {4.1.5g}** activities. The Customer Service function will also create, maintain and delete consumer accounts. It will **bill {4.1.5a}** and collect **payment {4.1.5c}** from consumers for the utilization of archive system resources.

* The **Data Engineering** function develops **submission agreements with data producers {4.1.2c},** supports ~~all~~ the Ingest Review function~~s~~ and is responsible for developing and maintaining the archive system data standards. The data submission formats and procedures must be clearly documented in the archive's data submission manual and the deliverables must be identified by the producer in the submission agreement. It will also develop **policies {4.1.3h}** for Archival Storage hierarchy management and migration policies to assure that archive storage formats do not become obsolete~~.~~

* The **System Engineering** function continuously monitors the functionality of the entire archive system and prepares recommendations and plans for system evolution.

* The **Management Interaction** function receives and carries out Management policies. These policies include the **OAIS charter {4.1.5g}**, **scope {4.1.5h}**, **resource utilization guidelines {4.1.5i}**, and **pricing policies (3.1.5j}**. It also provides OAIS **performance information {4.1.5k}** to Management.

4.1.6 ACCESS

The functions of the Access entity are detailed below:

* The **Provide Access Session** function provides a user interface to the information holdings of the archive. This interface will normally be via computer network or dial up link to an on line service, but might also be implemented in the form of a printed catalog, or fax-back type service. This function also provides a hierarchy of security controls depending on the needs of the archive system. These include establishing firewalls to prevent communication outside an area, electronic signatures and authorization procedures, restricting access to certain network domains, and assignment of user names and passwords. Any combination of these procedures may be needed in certain archive scenarios.

* The **Provide~~epare~~ Finding Aids** function provides tools and products which provide an overview of AIUs and AICs available in the archive system. Finding aids include summary versions of products which can be quickly viewed such as thumbnails images, or abstracts of documents. This function also generates **requests** for specialized queries or processing functions to be carried out by Dissemination to produce new representations of the data objects to extend the retrieval capabilities of the Data Management function (e.g., data mining).

* The **Accept Dissemination Request** function accept a **dissemination request {4.1.6b}** from a user, insure its validity, verify that all required information has been provided, and prepare the request for execution by the Dissemination entity. The dissemination request may be a **subscription request {4.1.6c}** or **adhoc request {4.1.6d}**. This function will provide the Consumer the opportunity to review

and correct the information in the request. It will also provides the capability to request special processing of data prior to dissemination via the process data function.

4.1.7 DISSEMINATION

The functions of the Dissemination entity are detailed below:

* The **Receive Dissemination Request** function accepts a **dissemination request {4.1.7a}**. A unique order number/identification is assigned to each accepted dissemination request. For each request accepted, this function adds an entry to the pending orders reflecting that this accepted order has not yet been filled. All order information is verified, and if any errors or unusual conditions are found the Consumer will be notified via an **error message {4.1.7k}**.

* The **Generate DIP** function accepts a dissemination request, validates the request using the **package descriptors {4.1.7j}**, **retrieves the data {4.1.7m}** from Archival Storage and moves a copy of the data to a staging area for further processing. This function also transmits a **report request {4.1.4b}** to Data Management. This function accepts and validates the commands, stores the validated selection parameters (field values) from the command, and then performs the necessary retrieval operations to find and access the requested data. It also accesses Data Management to obtain consumer information such as consumer's name, address, account number, preferred distribution method or media, and other consumer-oriented information. If special processing is required the generate DIP, the function provides a processing request {internal flow} to the Process Data function itemizing the data object in the staging area which require processing and the processes to be applied. The processing function will provide a notice of completed processing {internal flow} and identify output data objects in the staging area.

* The **Process Data** function receives a processing request from the Generate DIP function. It accesses data objects in staging storage and applies requested processes. The types of operations which may be carried out include statistical functions, sub-sampling in temporal or spatial dimensions, conversions between different data types or output formats, and other specialized processing (e.g. image processing). The Process Data function provides a notice of completed processing {internal flow} to the Generate DIP function upon completion of processing.

* The **Delivery** function handles both on-line and off-line deliveries of DIPs to consumers. For on-line delivery it accepts a DIP and prepares it for distribution in real time via Access **{4.1.7b}** over communication links. It identifies the intended recipient, the transmission procedures requested, and the DIP in the staging area to be transmitted. For off-line delivery it retrieves the DIP from the generate~~prepare~~ DIP function, prepares packing lists ~~, bills of lading~~ and other shipping records, and then ships the **DIP {4.1.7f}**. When the DIP has been shipped, a **notice of processed order {4.1.7e}** is returned to Administration. This function also calculates and records **billing information {4.1.7i}** for delivered orders and supplies them to the accounting function in Administration.

* The **Monitor Requests** function will track a dissemination request from inception to receipt of data by the consumer. It sends an **order confirmation {4.1.7h}** to the consumer, and notifies a consumer when his order has been executed. The confirmation fully identifies the order including order number, date of order, date of execution, identity of data requested, and method of distribution. The Monitor Requests function tracks every order from inception to delivery confirmation. Operations personnel are able to query the pending order file to determine the number and content of each unfilled order. As each order is filled (either by automated or manual means), it is removed from the pending order file. ~~The Monitor Requests function also has the ability to execute a standing order (a standard query and report procedure), based on elapsed time or some other trigger function.~~

4.1.8 DATA FLOW AND CONTEXT DIAGRAMS

The flow of data items among the OAIS functional entities is diagrammed in this section. **Figure 4-2** shows the more significant data flows. The flows associated with the Administration are generally support background activities of the other entities. To avoid complication of Figure 4-2, these background flows are illustrated in the context diagrams of **Figure 4-3** .

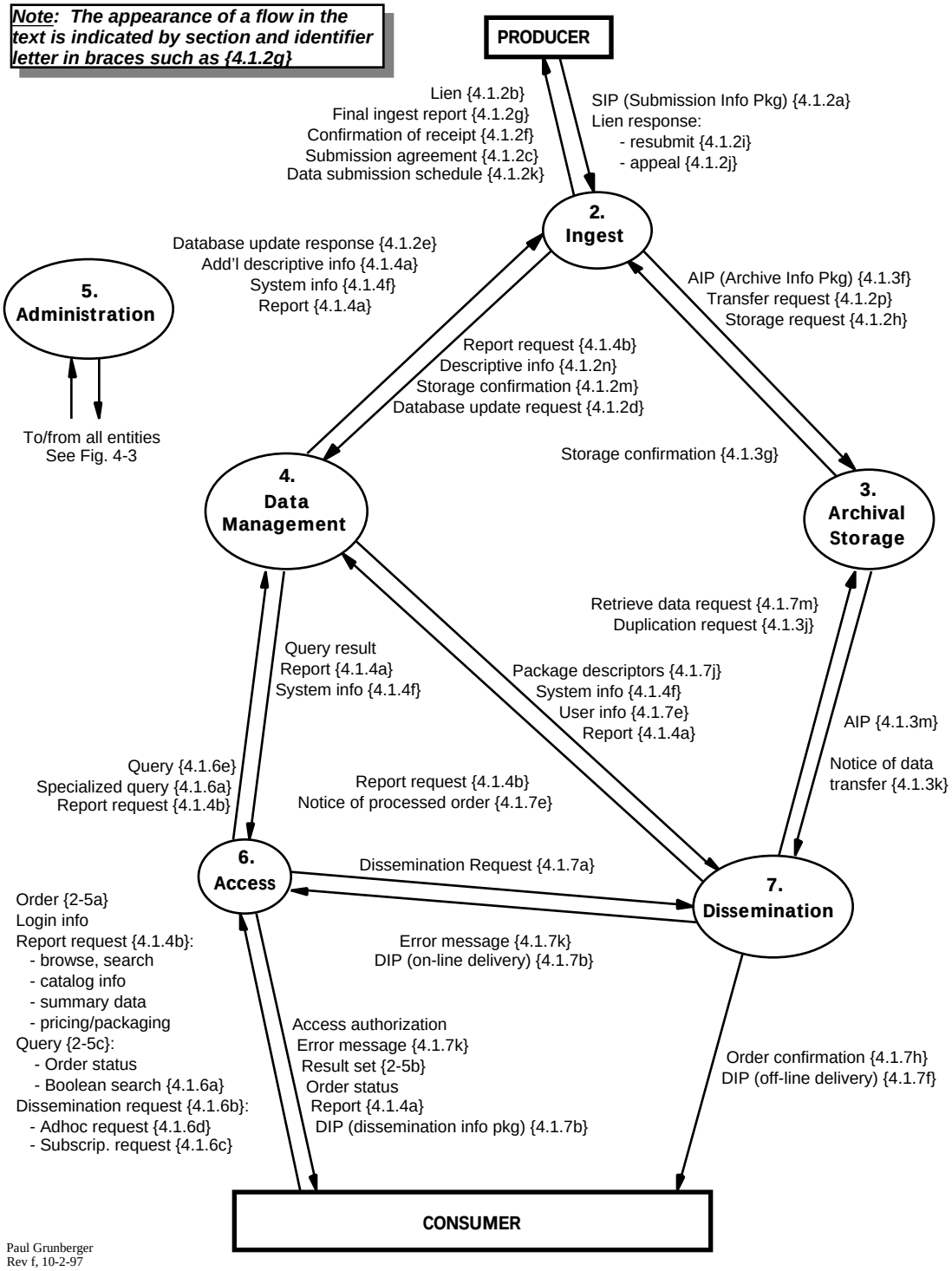

Figure 4-2. OAIS Data Flow Diagram

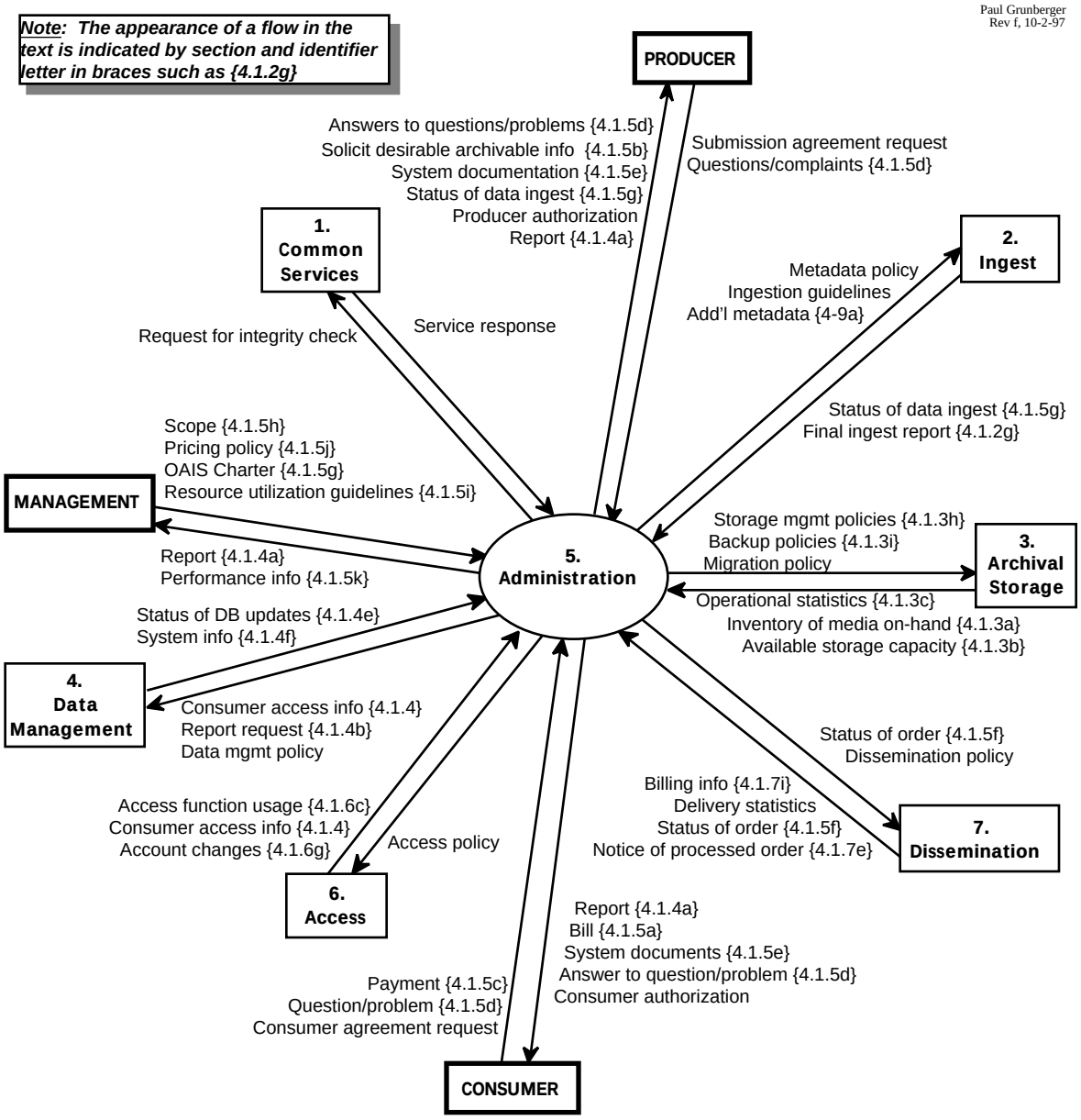

Figure 4-3. Administration Context